

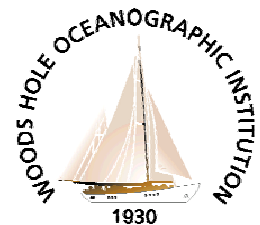
U.S. JGOFS Data Management *a retrospective*

Cyndy Chandler

U.S. JGOFS Data Management Office

25 January 2005

NACP Data Management Planning Workshop
New Orleans, LA





26 January 2005

blizzard of '05

Topics in today's presentation:

■ Introduction

- What is U.S. JGOFS?
- What is the DMO?

■ Lessons Learned

- U.S. JGOFS data server
 - JGOFS d-DBMS and user interface
 - Live Access Server

What is U.S. JGOFS?

U.S. Joint Global Ocean Flux Study

- part of multinational JGOFS
- U.S. Global Change Research Program (US GCRP), Scientific Committee on Oceanic Research (SCOR), and International Geosphere-Biosphere Programme (IGBP)
- long term (U.S. 1989-2005)
- multidisciplinary (bio, chem, PO, geology)
- process studies (U.S. 1989-1998), time-series, global surveys, synthesis and modeling, data management
- investigate ocean carbon flux

U.S. JGOFS

Data Management Office (DMO)

- formed in 1988 specifically to meet needs of U.S. JGOFS
- assist PIs to submit their data to DMO
- ongoing quality control of data
- develop and maintain simple, reliable interface to program data
- provide timely, easy access to project results
- collaborate with other program DMO
- publish U.S. JGOFS data reports
- plan final archive of U.S. JGOFS information

Basic Principles of Data Management

from 1988 JGOFS Working Group on Data Management

- scientists will generate data in a format useful for their needs
- oceanographic data sets are best organized in terms of metadata (temporal and geographical)
- data managers should avoid use of coded data values
- users should be able to obtain all the data they require from one source and in a consistent format
- data interchange formats should be designed for the convenience of scientific users

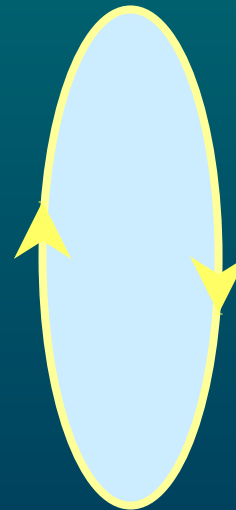
Additional Guidelines

- metadata is critical and therefore mandatory
- data managers should maintain awareness of emerging standards and strive for compliance
- whatever interface is used to provide access to the data, it's still important to provide subset and download capability
- data management systems must be dynamic - balancing tension between existing and new technologies

Basic Components of a Data Management System

first four components are active throughout the program

- data acquisition
 - from variety of sources
- quality assurance
- data publication
- synthesis and modeling



-
- archive

lesson defined . . .

1 : a passage from sacred writings read in a service of worship

2 a : a piece of instruction **b** : a reading or exercise to be studied by a pupil **c** : a division of a course of instruction

3 a : something learned by study or experience
b : an instructive example

4 : an edifying example or experience

5 : a reprimand

Lessons Learned . . .

The first lesson learned . . .

is a meta lesson . . .

All the lessons learned take on enhanced meaning when applied to science programs of increasing size and complexity.

Lessons Learned . . .

Lessons Learned . . .

- technology is good; people are more important
 - ✓ diverse range of expertise and personalities
 - ❖ designers, programmers, data managers
 - ✓ someone with authority to set overall vision
- ✓ qualified staff to make effective use of technology
- ✓ guidance from advisory committee which includes active investigators

U.S. JGOFS DMO personnel

- ❖ David Glover (director)
- ❖ Cyndy Chandler (manager)
- ❖ Jeff Dusenberry (data specialist)

- ❖ Previous staff members
 - Christine Hammond (manager)
 - George Heimerdinger (data specialist)
 - David Schneider (data specialist)

Lessons Learned . . .

- develop a data policy, publicize and follow it
 - ✓ guidance from steering committee
 - ✓ consent from participating investigators
 - ✓ reiterate at conferences and workshops
 - ✓ encourage compliance
- ✓ agree on method of enforcement (used as last resort)

Lessons Learned . . .

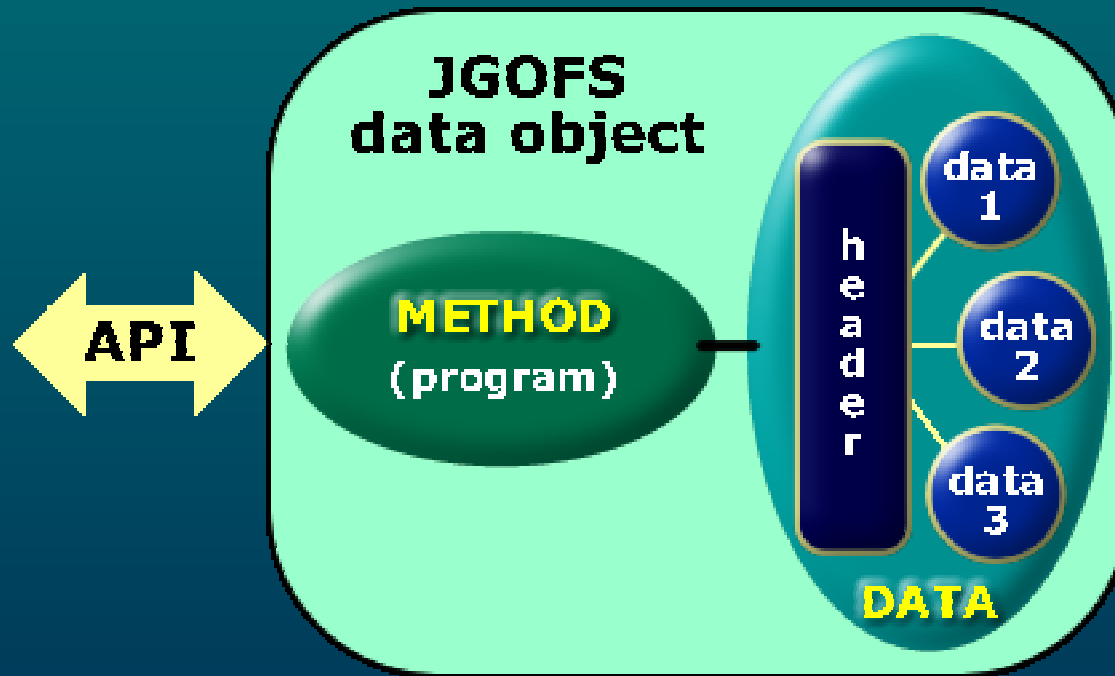
- establish protocols at program start with mechanism for adaptation when necessary
 - ✓ sampling methodologies
 - ✓ naming conventions
 - ❖ parameter dictionary, controlled vocabulary
 - ❖ XML schema, thesauri, ontologies
 - ✓ units of measurement

Lessons Learned . . .

- facilitate contribution of data to collection
 - ✓ compile an inventory of expected results
 - ✓ publish and maintain the inventory
 - ✓ remind investigators of opportunities to contribute data and results to the growing inventory
 - ✓ review procedure at conferences and workshops
 - ✓ accept all formats of data
 - ✓ work with investigators to complete metadata records

JGOFS distributed database management system

JGOFS object = method + data



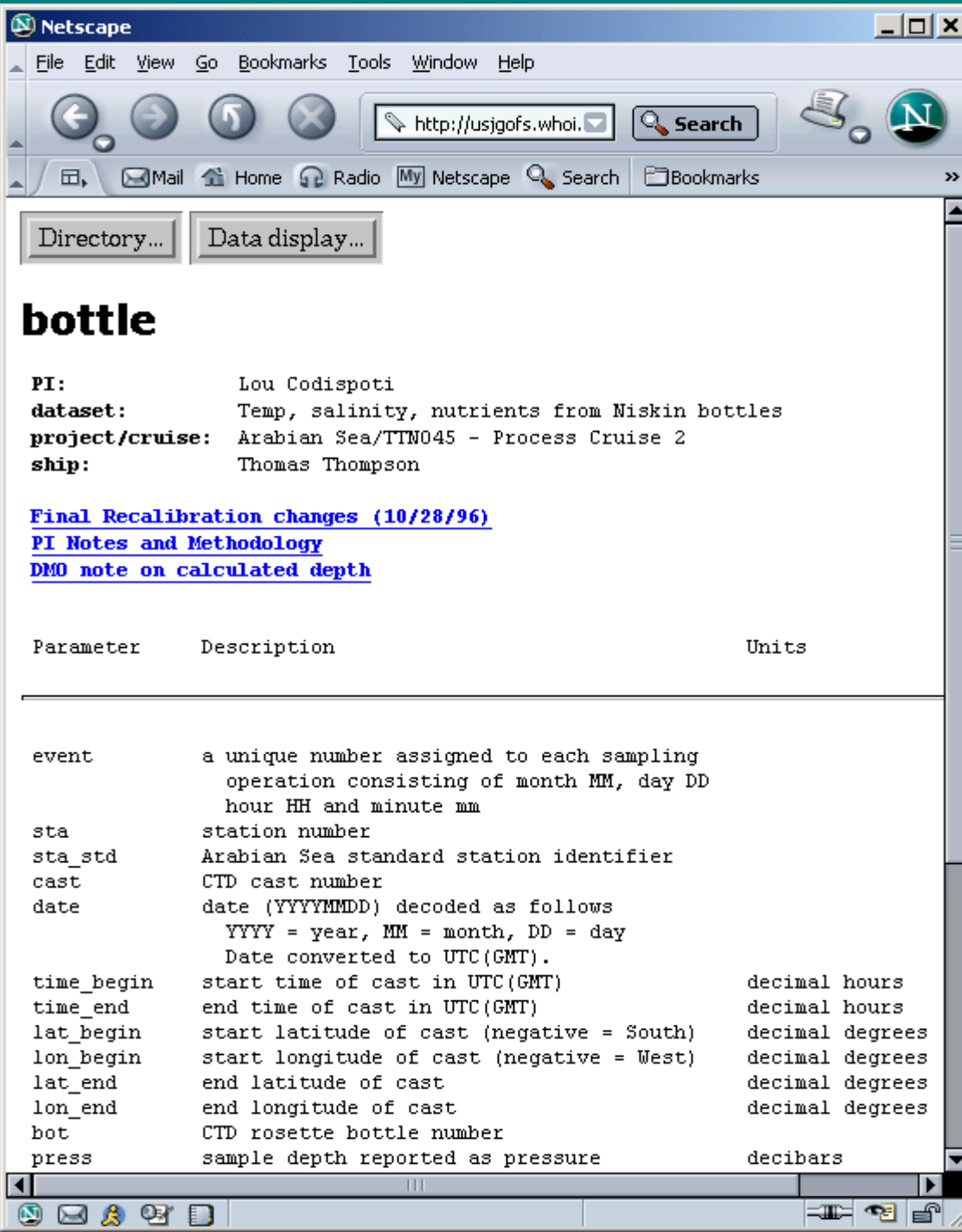
U.S. JGOFS DMO accepted any format data from the field study investigators and used methods to locate and translate the data objects

Lessons Learned . . .

➤ metadata is of critical importance

- ❖ accurate, complete, available with data
- ❖ monitor emerging standards
- ❖ define minimal metadata requirements
- ❖ standards-compliant solutions where possible
- ❖ complete metadata record enables reuse of data

➤ metadata assembly is time consuming, but is the key to enabling secondary reuse of data



metadata

Lessons Learned . . .

- quality assurance is an ongoing process
 - ❖ intense QA process during initial acquisition and ingestion into data system
 - ❖ problems discovered as data are utilized by others
 - ◇ insufficient or inaccurate metadata
 - ❖ process of data product synthesis becomes a valuable diagnostic tool for improving data quality

Lessons Learned . . .

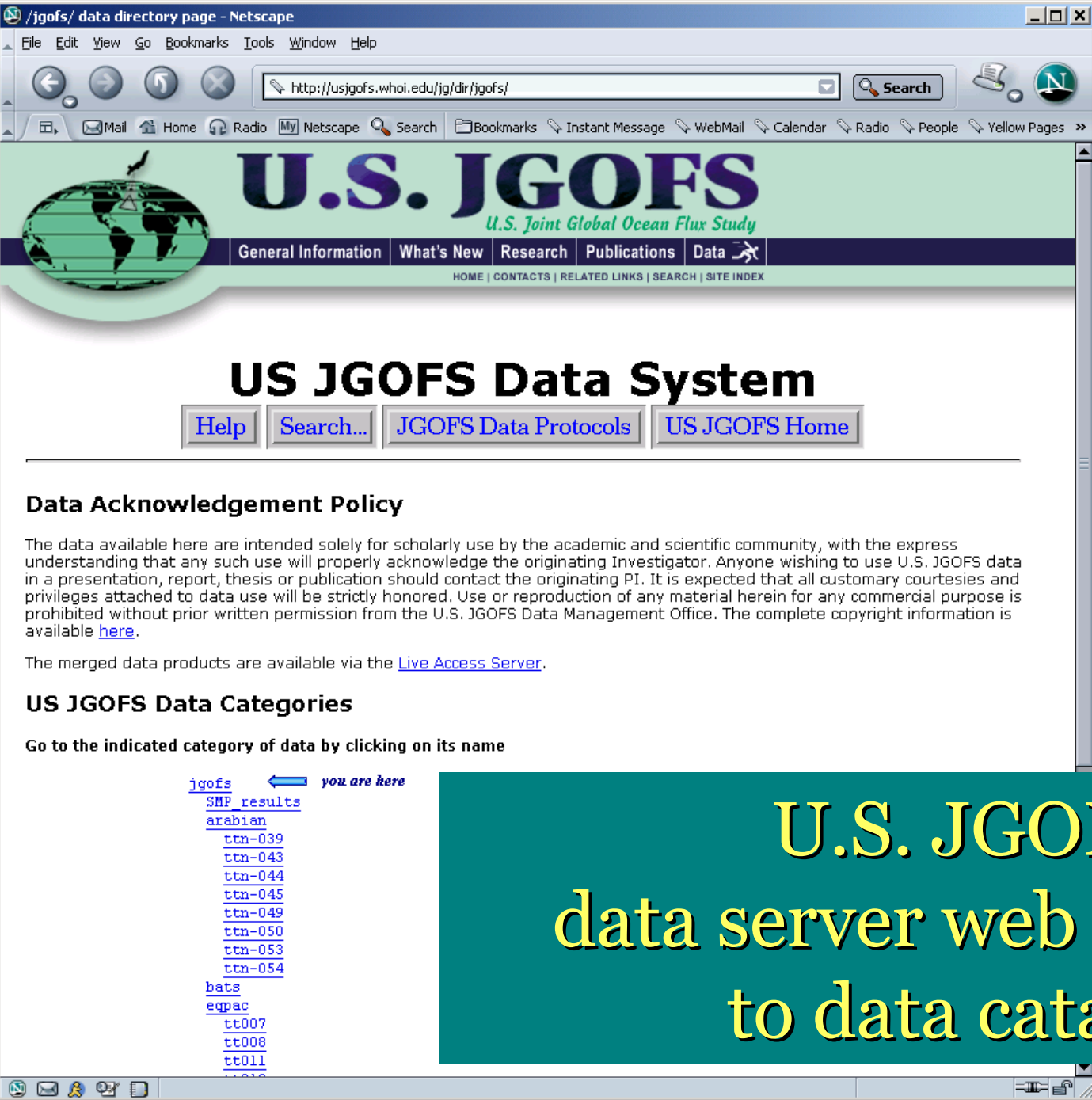
- begin synthesis early
 - ✓ do not wait until all the data has been collected
 - ✓ synthesized products greatly enhance the data collection

Lessons Learned . . .

- provide timely, easy access to project results
- develop and maintain simple, reliable interface to data collection
 - ❖ single interface to entire data collection
 - ❖ balance tension between new innovative technologies and existing stable implementations
 - ❖ if an interface is broken, it doesn't matter how great the original concept was

U.S. JGOFS Data Server

- JGOFS distributed database management system (d-DBMS) used for field data
- Live Access Server (LAS) used for gridded, synthesis and model results



U.S. JGOFS
data server web interface
to data catalog



JGOFS Distributed Database Management System (d-DBMS)

- distributed, object-oriented system
- originally developed by Glenn Flierl, James Bishop, Satish Paranjpe, David Glover
- supports multidisciplinary, multi-institutional data acquisition project
- multiple data storage formats and locations
- data interpreted by 'methods'

/jgofs/arabian/ttn-045/bottle ---- Level 1

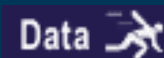
Directory... Documentation Plotting and Other Operations...

Level 0 Next level Flat listing

```
# version May 8, 2000
# Lou Codispoti
# Hydrographic data (temp, salinity and nutrients)
# Thomas Thompson TTN-045, Process Cruise 2 to Arabian Sea
#
=====
event      sta  sta_std  cast  date      time_begin  time_end  lat_begin  lon_begin  lat_end  lon_end
-----
03142214  1    N1      1     19950314  22.2461    22.9030   22.38423   59.88413   23.38506  58.87938
=====
bot  press  depth  temp  sal_ctd  sal_bot  O2_ml_L  O2_umol_kg  O2_umol_L  NO3  P04  Si04  NO2
-----
24   1.9    1.9    23.792 36.594  36.593  nd      nd          nd          5.07  0.79  3.4   0.49
23   1.8    1.8    23.788 36.594  36.589  4.254   185.33     189.98     5.12  0.81  3.4   0.49
22   12.0   11.9   23.793 36.594  36.590  nd      nd          nd          5.25  0.81  3.4   0.51
21   12.0   11.9   23.792 36.594  36.589  4.214   183.58     188.19     5.30  0.81  3.4   0.52
20   22.0   21.9   23.791 36.594  nd      nd      nd      nd      nd      nd      nd      nd
19   22.3   22.2   23.766 36.594  36.593  4.084   177.92     182.39     6.12  0.85  3.3   0.64
18   32.1   31.9   23.755 36.606  36.602  nd      nd          nd          5.89  0.84  3.2   0.66
17   32.0   31.8   23.755 36.606  36.601  4.157   181.10     185.65     5.89  0.84  3.2   0.65
16   42.1   41.8   23.659 36.592  36.588  4.078   177.66     182.12     6.69  0.88  3.3   0.38
15   42.2   41.9   23.660 36.592  36.595  4.083   177.88     182.34     6.76  0.89  3.3   0.36
14   52.1   51.8   23.203 36.500  36.544  nd      nd          nd          9.81  1.08  4.8   0.14
13   52.3   52.0   23.302 36.522  36.545  3.421   149.04     152.78     9.92  1.08  4.8   0.14
12   62.4   62.0   22.733 36.460  36.463  2.660   115.88     118.79     13.05 1.35  7.8   0.10
11   62.4   62.0   22.738 36.460  36.464  2.639   114.96     117.86     13.15 1.36  7.8   0.11
10   72.3   71.8   21.868 36.336  36.332  nd      nd          nd          18.49 1.77  11.4  0.09
9    72.1   71.6   21.867 36.336  36.328  1.435   62.51     64.09     18.64 1.79  11.4  0.10
8    102.4  101.7  20.761 36.186  36.185  0.531   23.13     23.71     22.43 2.15  15.4  0.06
7    102.3  101.6  20.760 36.186  36.185  0.525   22.87     23.45     22.47 2.23  15.6  0.06
6    132.4  131.5  19.937 36.142  36.134  0.220   9.58      9.83      23.40 2.38  18.3  0.04
5    132.4  131.5  19.934 36.136  nd      nd      nd      nd      nd      nd      nd      nd
4    162.5  161.4  18.967 36.220  36.223  0.127   5.53      5.67      22.93 2.39  21.1  0.03
3    162.4  161.3  18.965 36.218  36.222  0.130   5.66      5.81      22.92 2.29  20.9  0.03
2    202.8  201.4  18.193 36.356  nd      nd      nd      nd      nd      nd      nd      nd
1    253.2  251.5  17.428 36.404  36.414  0.170   7.40      7.59      22.32 2.35  24.8  0.05
=====
```

Data Listing

- ✓ select dataset
- ✓ select variable (columns)
- ✓ select range (rows)
- ✓ view metadata



subset, plot, download data

The screenshot shows a Netscape browser window with the title 'Plotting and Other Operations menu - Netscape'. The address bar contains 'http://optserv1.whoi.edu/jg/otherc'. The main content area is titled 'Plotting and Other Operations Menu' and displays the current object path: '//usjgofs.whoi.edu/jgofs/arabian/ttn-045/bottle'. Below this, there are four main categories of operations, each with a list of links and a 'Help' button:

- Listing and downloading data
 - [List at this level](#)
 - [Other data listing formats](#)
 - [Matlab file format](#) of all data at this level and further in. [Help](#)
 - [Download utility](#) [Help](#)
- Manipulating data
 - [Math operations](#) for calculating values from existing parameters. [Help](#)
 - [Join 2 objects](#) having at least 1 parameter in common. [Help](#)
 - [Statistics](#)
- Plotting data
 - [Simple X-Y plot](#)
 - [Mapping of data locations](#) (JGOFS software)
- [Subsetting data](#) [Help](#)

At the bottom of the page, it says 'Version: April 1, 2004'. The browser's status bar at the very bottom shows various icons for navigation and utility.

Live Access Server Interface

provides
access to
synthesis
and model
results

Live Access to U.S. JGOFS Data - Netscape

File Edit View Go Bookmarks Tools Window Help

http://usjgofs.whoi.edu:8089/las/servlets/dataset Search

Mail Home Radio My Netscape Search Bookmarks Instant Message WebMail Calendar Radio People

 **U.S. JGOFS**
U.S. Joint Global Ocean Flux Study

General Information What's New Research Publications Data 

HOME | CONTACTS | RELATED LINKS | SEARCH | SITE INDEX

Live Access to U.S. JGOFS Data Search: Go

single data set compare two

Datasets

Variables

Constraints

Output

Output Options

Previous Output

Define variable

About

LAS UI Version 6.2.1

Datasets

Click on a dataset to continue or an  for information about a dataset. [Help](#)

Select dataset:

[Climatology](#)

[In situ Data Synthesis](#)

[Model Results](#)

[Ocean Color](#)

[Synthesis Projects](#)

Live Access Server ~ LAS

- LAS Development Team (original)
 - Steve Hankin
 - Jon Callahan
 - Joe Sirott
- located at UW JISAO/NOAA-PMEL
University of Washington's Joint Institute for the
Study of the Atmosphere and Ocean and NOAA
Pacific Marine Environmental Laboratory

Live Access Server Interface

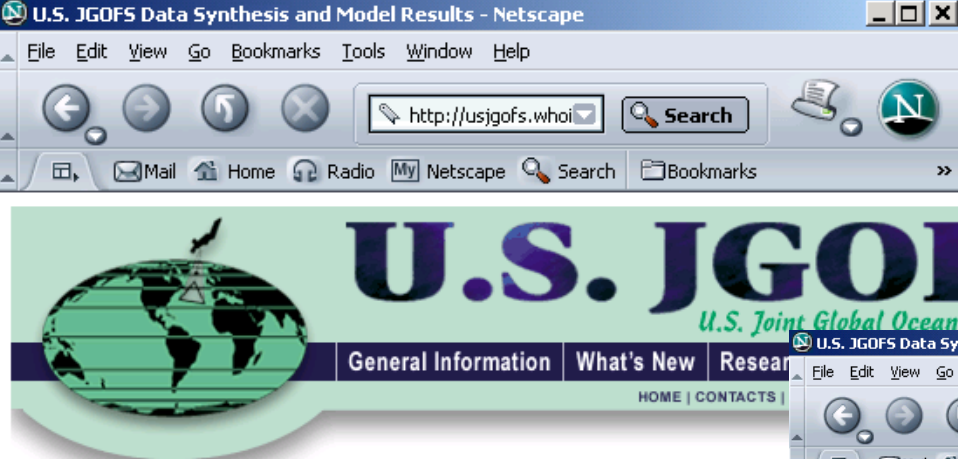
- configurable Web server
- data and metadata interface
- provides access to geo-referenced scientific data
- presents distributed data sets as a unified virtual data base (DODS/OPeNDAP)
- uses Ferret as the default visualization application
- visualize data with on-the-fly graphics
- request custom subsets of variables in a variety of file formats
- access background reference material (metadata)
- compare variables from distributed sources

LAS enables a data server to ...

- unify access to multiple types of data in a single interface
- create thematic data collections from distributed data sources
- offer derived products on the fly
- offer variety of visualization styles
 - customized for the data

U.S. JGOFS LAS

- MySQL database of netCDF and JGOFS format data objects
- interface to project data and metadata
- data sub-selection (selections, projections)
- multi-variable support
- gridded vs. in-situ data differencing
- multiple views
 - (property-property, depth horizon, cruise tracks, overplots)
- multiple products (ps, gif, text, NetCDF)



LAS v6

select dataset and variables

U.S. JGOFS Data Synthesis and Model Results

single data set

compare two

Datasets

Variables

Constraints

Output

Output Options

Previous Output

Define variable

About

LAS UI Version 6.3

Datasets > In situ Data Synthesis

Click on a dataset to continue c

Select dataset:

- U.S. JGOFS Arabian Sea CTD Data
- U.S. JGOFS Arabian Sea Niskin Bottle Data
- U.S. JGOFS Arabian Sea TM Bottle Data
- U.S. JGOFS Equatorial Pacific CTD Data
- U.S. JGOFS Equatorial Pacific Niskin Bottle D
- U.S. JGOFS North Atlantic CTD Data
- U.S. JGOFS North Atlantic GoFlo Bottle Data
- U.S. JGOFS North Atlantic Niskin and GoFlo E
- U.S. JGOFS North Atlantic Niskin Bottle Data
- U.S. JGOFS Southern Ocean CTD Data
- U.S. JGOFS Southern Ocean Niskin Bottle Da
- U.S. JGOFS Southern Ocean TM Bottle Data

U.S. JGOFS Data Synthesis and Model Results

single data set

compare two

Datasets

Variables

Constraints

Output

Output Options

Previous Output

Define variable

About

LAS UI Version 6.3

Datasets > In situ Data Synthesis > U.S. JGOFS Arabian Sea Niskin Bottle Data

Select a variable and then click **Next >** to proceed to the Constraints page. [Help](#)

Merged Data Product compiled from *in situ* U.S. JGOFS Arabian Sea Niskin Bottle Data

[Merged Product Metadata](#)

Dataset variable(s): [Reset](#) | [Select all](#) | [Unselect all](#)

PHYSICAL PROPERTIES

- ☐ salinity, water bottle sample, PSS-78 scale
- ☐ salinity, from CTD unit when water bottle tripped
- ☐ temperature, from CTD, IPTS-68
- ☐ depth of mixed layer, calculated based on a .1 deg. C change in temperature
- ☐ depth of mixed layer, calculated based on a .5 deg. C change in temperature
- ☐ depth of mixed layer, calculated based on a .03kg/meter^3 change in density from the surface, reported in decibars
- ☐ depth of mixed layer, based on a .05kg/meter^3 change in density from the surface, reported in decibars
- ☐ depth of mixed layer, calculated based on a .125kg/meter^3 change in density from the surface, reported in decibars
- ☐ depth of mixed layer, calculated based on a .25kg/meter^3 change in density from the surface, reported in decibars

[Next >](#)

LAS v6 constraints

- ✓ select dataset
- ✓ select variable
- set constraints
- select view (XYZT)
- select output type
- selections:
 - lat/lon
 - time
 - depth range

26 January 2005

Live Access to U.S. JGOFS Data - Netscape

File Edit View Go Bookmarks Tools Window Help

http://usjgofs.whoi.edu:8089/las/servlets/constrain?var=85 Search

Mail Home Radio My Netscape Search Bookmarks Instant Message WebMail Calendar Radio People

U.S. JGOFS

U.S. Joint Global Ocean Flux Study

General Information What's New Research Publications Data

HOME | CONTACTS | RELATED LINKS | SEARCH | SITE INDEX

Live Access to U.S. JGOFS Data

Search: Go

single data set **compare two**

Datasets
Variables
Constraints
Output
Output Options
Previous Output
Define variable
About

LAS UI Version 6.2.1

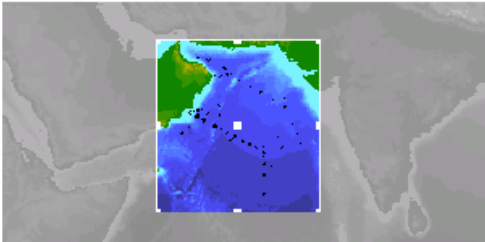
U.S. JGOFS Arabian Sea Niskin Bottle Data
Variable(s): **Ammonium**

Select your desired view (geometry of output) and output (type of product). Then set the 4-D region (lon-lat-depth-time) and any additional constraints. [Help](#)

Select view: Longitude-Latitude

Select output: Pie plot (GIF)

Select region: Arabian Sea Study Area [Go](#) [Don't use map applet](#)



26.0 N
54.0 E 71.0 E
8.0 N
[Zoom In](#) [Zoom Out](#)

Select time range: 01 Jan 1995 to 31 Dec 1995

Select depth range: 0 to 4900 100

Select Constraints:

Apply: ☐ 19-prime-butanoyloxyfucoxanthin <

Apply: ☐ Standard Station = A

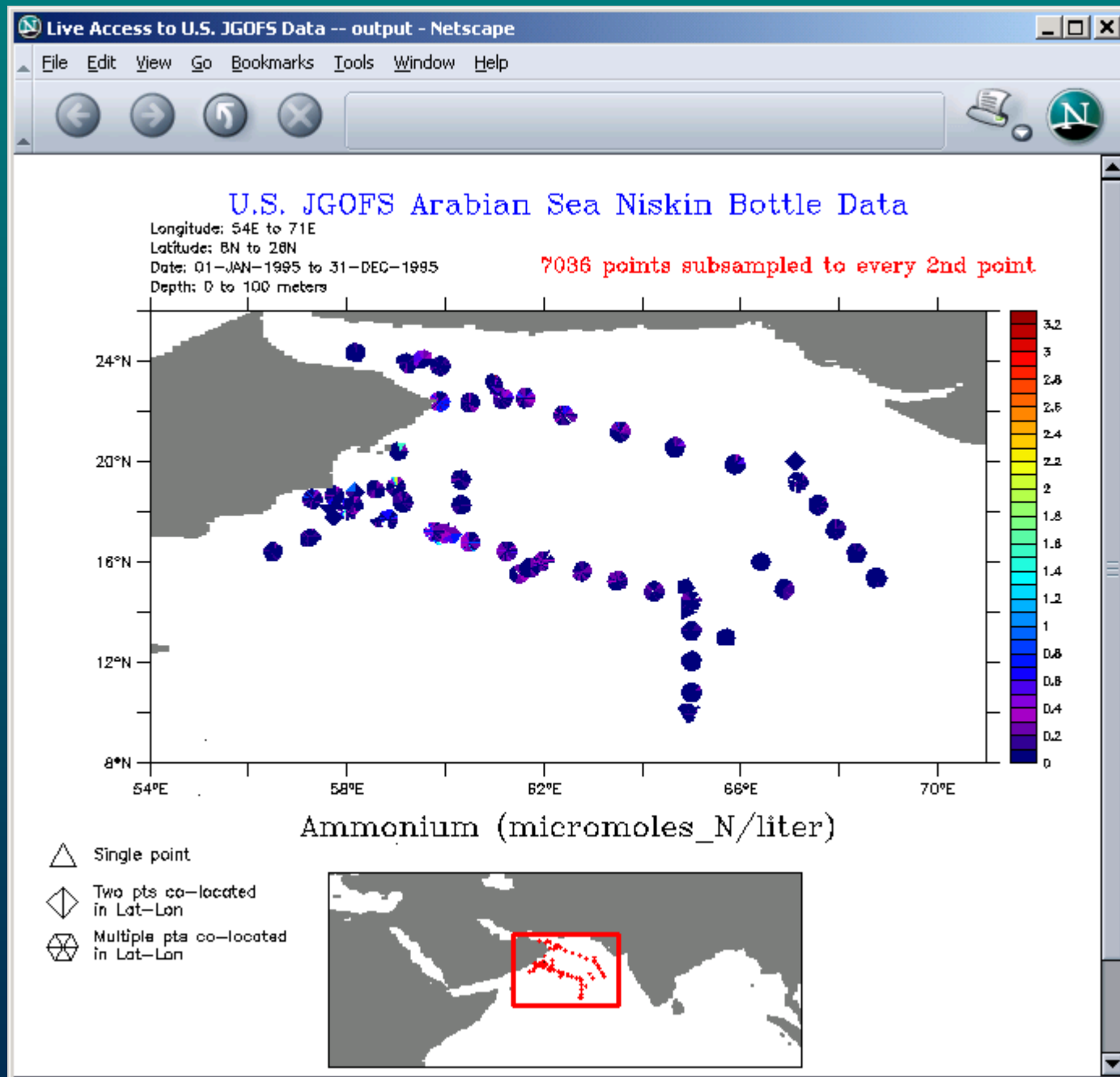
Apply: ☐ Cruise ID = PC1 TTN-043 (Jan-08:Feb-05)

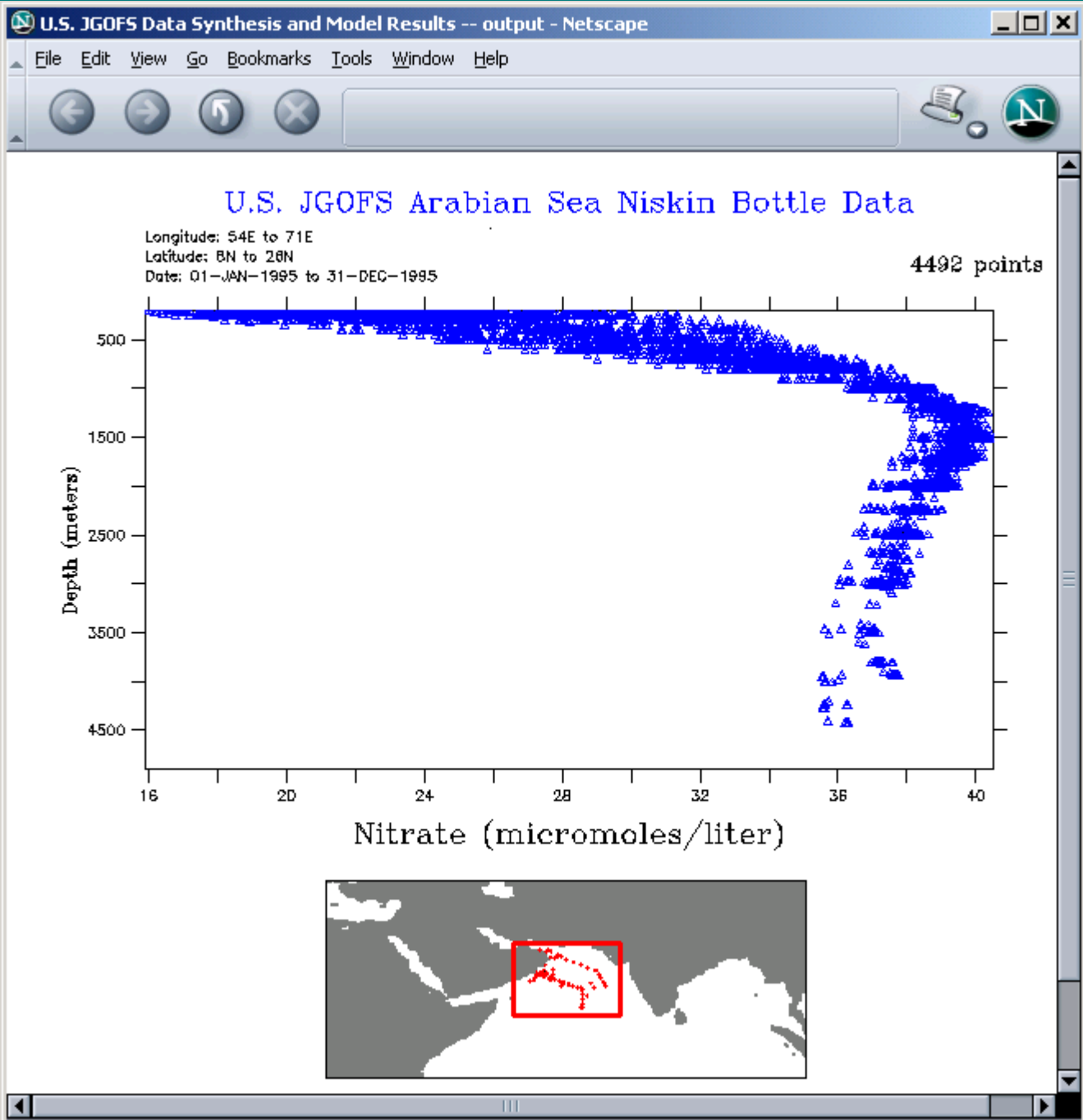
Applet map started

- ✓ select dataset
- ✓ select variable
- ✓ set constraints
- output

LAS

26 January 2005





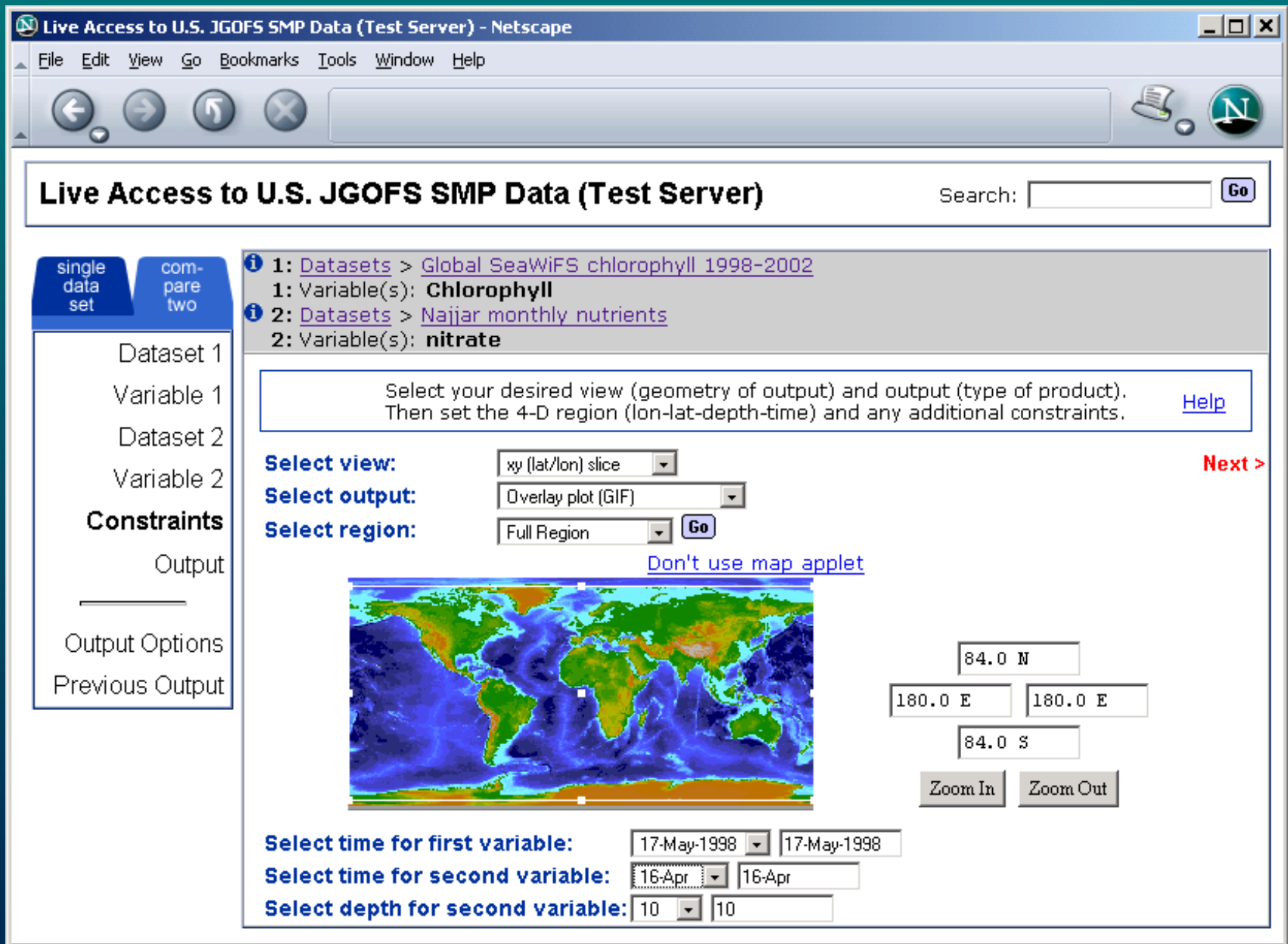
LAS
v6

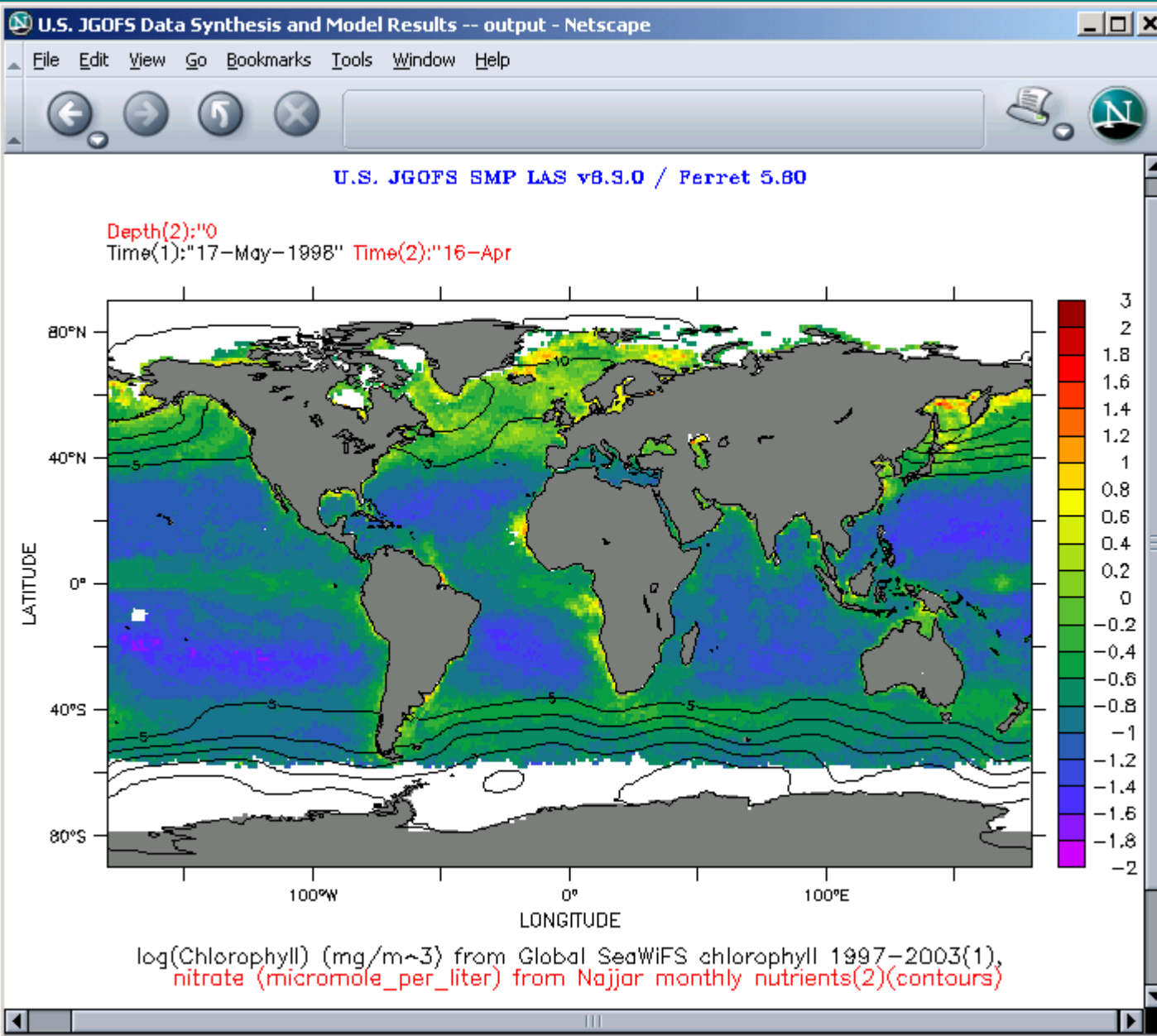
Property-
Depth

Arabian Sea
Nitrate

from merged
product
generated by
DMO from *in situ*
Niskin bottle data

Comparison Overlay Plot





chlorophyll
(May)
shaded

SeaWiFS data
contributed by:
Yoder and
Kennelly

surface
nitrate (April)
contours

Nutrient
Climatologies
contributed by:
Ray Najjar

U.S. JGOFS Data System Summary

- supports a variety of data formats
- OPeNDAP used to access data collection
- coupled metadata and data
- supports data subselection
- offers variety of products for download

Lessons Learned . . .

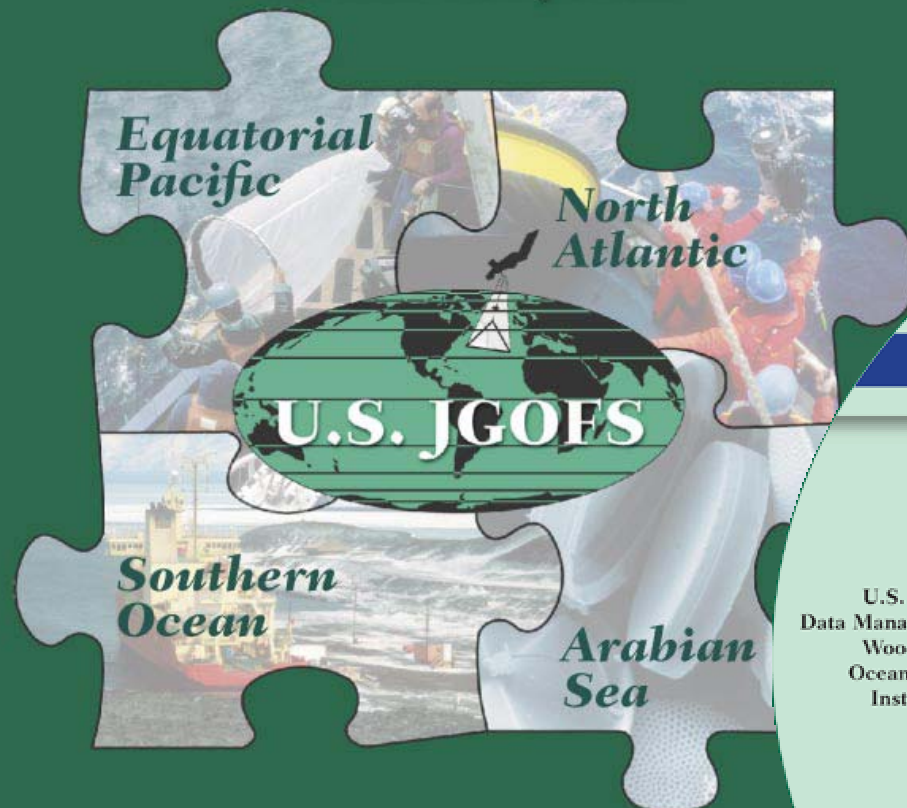
- encourage data managers to collaborate
 - ✓ other data managers within program
 - ❖ JGOFS DMTT
 - ✓ data managers from other programs
 - ❖ GLOBEC, LTER
 - ✓ program investigators and participants
 - ❖ attend conferences and workshops
 - ❖ offer data system tutorials

Lessons Learned . . .

- publish data reports
 - ✓ archive data in one place
 - ✓ easy access to project results
 - ✓ most complete and accurate form of database

**United States Joint Global Ocean Flux Study
Final Data Report, Volume 1
Process Study Data**

data reports published on
CD-ROM



Funded primarily by the U.S. National Science Foundation
with additional support from NOAA, NASA, DOE and ONR



Lessons Learned . . .

- plan early for final archive of program results
 - ✓ digital records
 - ✓ don't forget the boxes of stuff !



Lessons Learned . . .

- develop a data policy, publicize and follow it
- establish protocols at program start with mechanism for adaptation when necessary
- facilitate contribution of data to collection
- **metadata is of critical importance**
 - ❖ accurate, complete, available with data
- quality assurance is an ongoing process
- begin synthesis early

Lessons Learned . . .

- provide timely, easy access to project results
- develop and maintain simple, reliable interface to data collection
- encourage data managers to collaborate
- publish data reports
- plan early for final archive of program results

Challenges

➤ functioning amid the chaos

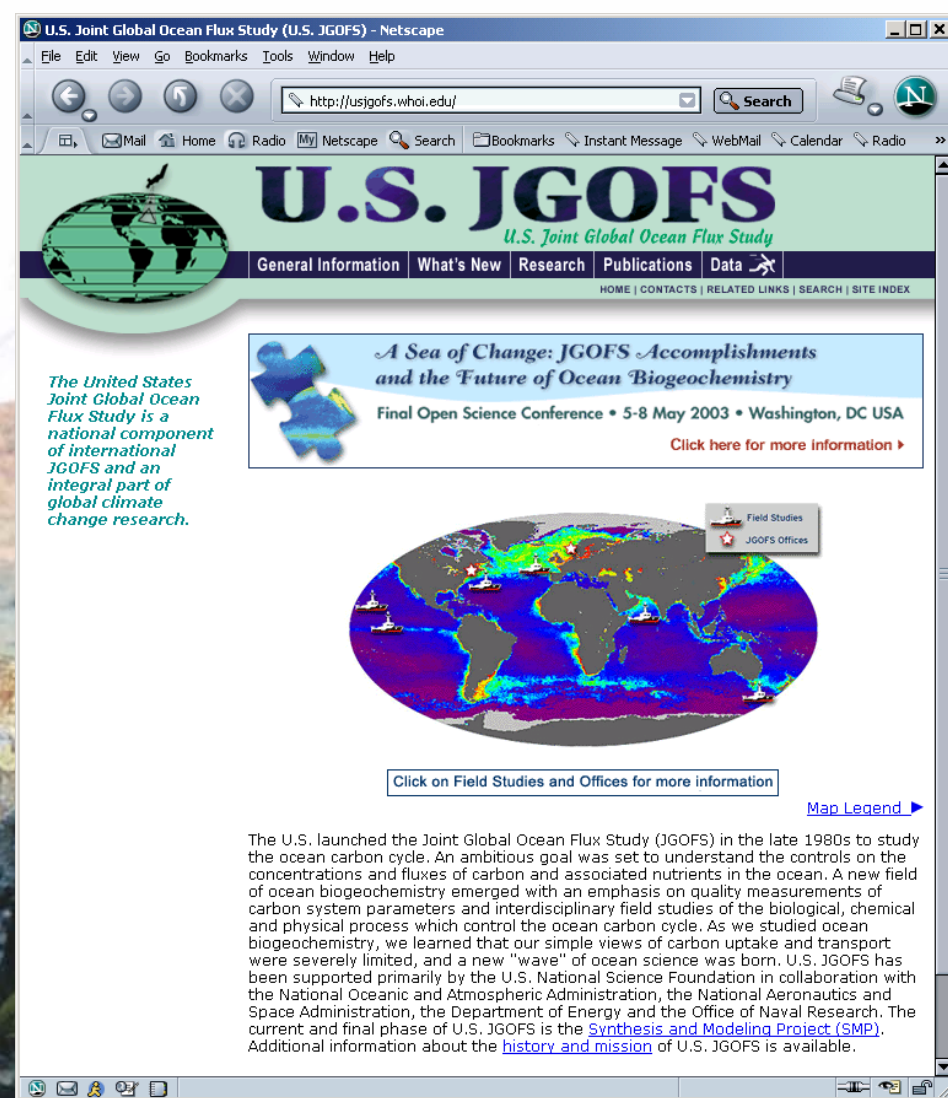
- ❖ maintaining a healthy data management system amid the chaos of rapidly changing information technology
- ❖ distinguishing between enabling and disruptive technologies

➤ data and results – what to preserve?

- ❖ raw and processed data, synthesized products, model code, inputs, results

➤ increasing volume and diversity

➤ long-term preservation of data



U.S. JGOFS web site
<http://usjgofs.who.edu>